



US006161094A

# United States Patent [19]

[11] **Patent Number:** **6,161,094**

**Adcock et al.**

[45] **Date of Patent:** **Dec. 12, 2000**

[54] **METHOD OF COMPARING UTTERANCES FOR SECURITY CONTROL**

*Primary Examiner*—David R. Hudspeth

*Assistant Examiner*—Susan Wieland

[75] Inventors: **Sherrie Adcock**, Dana Point, Calif.;  
**Gerald S. Schwartz**, Kirkland, Wash.

*Attorney, Agent, or Firm*—Gene Scott-Patent Law & Venture Group

[73] Assignee: **Ann Adcock Corporation**, Dana Point, Calif.

[57] **ABSTRACT**

[21] Appl. No.: **09/190,386**

A security method compares a present verbal utterance with a previously recorded verbal utterance by comparing frequency domain representations of the utterances, with multiple repeat utterances forming a basis for determining a variation in repetitious performance by an individual, and similar differences between enrollment and challenge utterances forming a basis for a similar analysis of variance between enrollment and challenge utterances. In one embodiment a set of enrollment data is searched by each challenge until either a match is made, indicating an action, possibly dependent upon the specific match, or no match is made indicating an abort.

[22] Filed: **Nov. 10, 1998**

**Related U.S. Application Data**

[63] Continuation-in-part of application No. 08/800,354, Feb. 14, 1997, Pat. No. 5,835,894.

[51] **Int. Cl.**<sup>7</sup> ..... **G10L 7/08**

[52] **U.S. Cl.** ..... **704/273; 704/274; 704/275; 704/254**

[58] **Field of Search** ..... **704/273, 274, 704/275, 254, 251**

[56] **References Cited**

**U.S. PATENT DOCUMENTS**

5,835,894 11/1998 Adcock et al. .... 704/273

**3 Claims, No Drawings**

## METHOD OF COMPARING UTTERANCES FOR SECURITY CONTROL

This application is a continuation-in-part of a previously filed application, Ser. No. 08/800,354 filed Feb. 14, 1997 now U.S. Pat. No. 5,835,894.

### BACKGROUND OF THE INVENTION

#### 1. Field of the Invention

This invention relates generally to electronic security methods which provide for modeling or otherwise comparing human features such as fingerprints, voice patterns, and retina patterns, in order to distinguish between individuals, and, more particularly, to a security method and protocol for modeling and comparing voice utterances to control the operation of a security device.

#### 2. Description of Related Art

The following art defines the present state of this field:

Parra, U.S. Pat. No. 5,313,556 describes the identity of an individual (known or unknown) that is determined by a sonic profile of sounds issued through his oral-nasal passages. The sounds are converted to digital electrical signals and produce a three domain format of frequency, amplitude and time samples to produce an array of peaks and valleys constituting the sonic profile of an individual. A source or library of sonic profiles in the same format of a known individual have an interrelationship including relative positions of said peaks and valleys of said sonic profile of the known individual with that of said unknown individual compared and a utilization signal is provided upon detecting or non-detecting a correlation between said sonic profiles.

Hair et al., U.S. Pat. No. 3,673,331 describes voice verification that is accomplished at a plurality of spaced apart facilities each having a plurality of terminals. Multiplexing structure interconnects the terminals through a communications link to a central processing station. Analog reproductions of voices transmitted from the terminals are converted into digital signals. The digital signals are transformed into the frequency domain at the central processing station. Predetermined features of the transformed signals are compared with stored predetermined features of each voice to be verified. A verify or non-verify signal is then transmitted to the particular terminal in response to the comparison of the predetermined features.

Waterbury, U.S. Pat. No. 3,896,266 describes a security card (which may be a credit card) according to the invention has recorded on it data identifying a personal and non-counterfeitable attribute, such as the voice characteristics, of the authorized holder of the card. A card utilization system according to the invention provides means for comparing the attribute as recorded by these data with the corresponding attribute of the person wishing to use the card, thereby substantially eliminating the possibility of unauthorized card utilization.

Muroi et al., U.S. Pat. No. 4,833,713 describes a voice or sound recognition system including a microphone for converting a voice into an electrical voice signal, a frequency analyzer for generating a voice pattern in the form of a time-frequency distribution, and a matching unit for matching the voice pattern with registered voice patterns.

Feix et al., U.S. Pat. No. 4,449,189 describes a method and an apparatus for identifying an individual through a combination of both speech and face recognition. The voice signature of an interrogated person uttering a key word into a microphone is compared in a pattern matcher with the

previously stored voice signature of a known person uttering the same key word to obtain a first similarity score. At the same time, when a key event in the utterance of the key word by the interrogated person occurs, a momentary image of that person's mouth region onto which a grid pattern has been projected is optically recorded and compared with the previously stored corresponding momentary image of the same known person to obtain a second similarity score.

The prior art teaches the comparing of voice signatures in time as well as frequency domain. However, the prior art does not teach a means for filtering such voice profiles by difference techniques. The present invention method fulfills these needs and provides further related advantages as described in the following summary.

### SUMMARY OF THE INVENTION

The present invention teaches certain benefits in methods which give rise to the objectives described below.

The present invention is a security method which compares a present verbal utterance with a previously recorded verbal utterance by comparing frequency domain representations of the present utterance, with previously recorded multiply repeated utterances of the same material, forming a basis for comparison. Instead of comparing the utterances in the time domain, as is conventional, the present method approaches the comparison by establishing energy content in a variety of cells in the frequency domain, and instead of focusing on the ability of an individual to repeat an utterance from one trial to the next, sometimes separated by days, weeks or even longer, the present approach focuses on the variability of the difference between multiple utterances of the same words or phrases. Specifically, the method attempts to determine if two sounds were produced by the same human voice in an attempt to discriminate between allowed and non-allowed personnel seeking to operate a secure device. Further, the method may be used to determine what command is being given by the individual, culling out the selected command from a library of such commands all uttered by the same individual.

The present method invention has as an important aspect, the discrimination between, and, or matching of a presently uttered verbal word or phrase with the same utterance stored in a library of such utterances. Another aspect of the present method is the achievement of high accuracy and fast results in discrimination and, or matching of verbal utterances by using a difference method for comparison. The present method is a non-obvious and highly effective procedure for extremely high speed comparison of large data sets against a challenge so as to provide the convenience, for instance, of verbal only challenges at a secure door used by a large number of individuals with separate verbal access codes, wherein the delay time for approval has been shown to be in the range of a few seconds. The method also results in an extraordinary level of discrimination between individuals while providing a high level of "forgiveness" for the normal changes in tone, timber and volume of the human voice from moment to moment and day to day. The discrimination capability of the present method is strong enough for use in military as well as industrial applications, and is inexpensive and simple to use so as to find application in residential use 51 commercial. A further aspect of the present method is the use of testing for minimum energy levels in a set of frequency ranges in the frequency domain, as well as testing for corresponding energy levels that surpass a selected energy level criterion.

Other features and advantages of the present invention will become apparent from the following more detailed description providing an understanding of the principles of the invention.

### DETAILED DESCRIPTION OF THE DISCLOSURE

The present invention provides a step-by-step method for comparing a verbal utterance of a speaker in the present (challenge utterance), with a previously recorded verbal utterance (enrollment utterance), to determine the validity of the speaker, i.e., if the challenge utterance is being produced by the same person as was the enrollment utterance. When validity is determined, an action is authorized, such as opening a lock, dialing a secret phone number, granting access to data or services, etc. The method comprises certain steps which are performed in sequence. In summary the steps are defined by, first, preparing an enrollment data from an utterance from one or more persons, next, challenging the enrollment data with a present utterance from one of those persons, and finally enabling the security action if the challenge utterance is accepted as being close enough to one of the enrollment data.

Preparing the enrollment data comprises the steps of, converting a verbal utterance, which we shall refer to an enrollment utterance, into a first electrical signal as by a microphone or other transducer. This electrical signal is transformed into a digital format. A fast Fourier transformation of this electrical signal is conducted to produce a frequency domain representation of the enrollment utterance. The Frequency domain representation is then divided into frames of time, e.g. 10 ms. Frames which show no energy content are deleted. A number of samples, represented by "M" are taken for each of N frequency channels in order to form an M by N sample enrollment matrix  $E_i$ . M and N are selected as integers of a magnitude necessary for the level of security desired with larger numbers providing greater security and vice-versa. This matrix provides cell samples  $M_i-N_i$ , where i represents an integer, which are characterized by a total energy content within each cell, i.e., a number. Next, the method provides for determining if at least X, an arbitrary number, of the M samples have a selected minimum energy content in at least Y, another arbitrary number, of the N frequency channels. If not, the enrollment utterance is repeated until the criteria X and Y are satisfied, or if not, after several tries, the process of enrollment is aborted. This usually would only happen if the speaker is not able to provide enough volume in his or her speech to enable production of the minimum energy criterion or if the input is not a human voice so that a necessary spectral content is lacking.

When the enrollment is successful, the process is repeated until a selected number of the sample matrices, herein designated as  $E_i$ , have been formed and stored. We shall refer to these matrices as  $E_1, E_2, E_3, \dots$ . Typically, we shall require at least three of these matrices in order to be satisfied with a successful enrollment of a voice utterance.

Next, the method requires forming the difference between each pair of the enrollment matrices  $E_i$ , as  $(E_1-E_2), (E_1-E_3), (E_2-E_3), \dots$ . Algorithms are applied such that each individual frame is compared with each other frame and is allowed to slip any number of frames, forward or backward in order to obtain a best match. Thus, we have determined a variance or variability characteristics of the individual speaker in forming the enrollment utterances.

Finally, we form the sum, S, of the differences between each of the pairs of the matrices  $E_i$ , as  $S=[(E_1-E_2)+(E_1-E_3)+(E_2-E_3)+\dots]$ , and accept S as an enrollment data only if  $S < EB$ , EB being a selected first enrollment criterion called the "enrollment basis," and also only if each of the differences between each pair of the matrices  $E_i$  is greater than

ES, ES being a selected second enrollment criterion called "enrollment span." The first part of this step is critical to the viability of the present method in that EB determines that the total energy in the differences,  $E_1-E_2$  for example, must be less than the EB criterion or else we must consider that the variances in repetitive spoken utterances of the same material, for this individual, is too great to be useful in the present method. This is clearly a novel step, with respect to the prior art and is critical to the determination of correlation between an utterance spoken at one point in time relative to the same utterance spoken at another point in time. The second part of this step is also critical to the viability of the present method in that ES assures us that a recording is not being used as an impostor, i.e., we can be sure that the differences in the matrices are at least as great as would be normally expected from a human voice. When a recording is used for each of the enrollments, we find that variability between them is less than is produced by the human voice.

Challenging the enrollment data comprising the steps of first, converting a challenge verbal utterance into a second electrical signal as by a microphone or other transducer. This electrical signal is converted into a digital format. The digital signal is used to perform a second fast Fourier transformation of the represented electrical signal to produce a frequency domain representation of the challenge utterance. The Frequency spectrum is then divided into frames of time, e.g. 10 ms. Frames which show no energy content are deleted. Taking M samples of the second determining if at least X of the M samples have a selected minimum energy content in at least Y of the N frequency channels. It is clear that these steps in the challenge are identical to the corresponding steps in the enrollment procedure. These challenge steps are repeating if the later test for minimum energy fails, and the challenge is aborted after a selected number of such repeats on the basis that an impostor is trying to crack the system and such an impostor should not be given as many tries as he/she would desire.

Next the difference between C and  $D_i$  for each of the matrices  $D_i$  as  $(C-D_1), (C-D_2), (C-D_3), \dots$  is formed. Algorithms are applied such that the individual frame being compared to any other frame  $(C-D_1)$  can slip time by an arbitrary number of frames, forward or backward in order to obtain the best match, which is formed. Then the sum, S, of the differences between each of the pairs C and  $D_i$ , as  $S'=[(C-D_1)+(C-D_2)+(C-D_3)+\dots]$  is formed. These steps correspond to similar steps in the enrollment process. S' is accepted as a valid challenge if  $S' < VB$ , VB being a selected first verification criterion and also if each said difference between each of the pairs C and  $D_i$  is greater than VS, VS being a selected second verification criterion. Both VB and VS play corresponding roles to EB and ES in the enrollment procedure and are used for the same reason. When these criteria are met the challenge verbal utterance is accepted and the requisite security step is enabled.

We have described the present method when used for an application where only one challenger is expected. In this case only one enrollment need be made. The objective in such a system is to assure that only the one enrolled individual can access the system by successfully challenging the method. All others, being impostors, must not be able to successfully challenge the method so that they are locked out of any security system using the method.

It is clear that such a system may use the method for a plurality of users at the same time when an enrollment is completed for each user. In this case, a challenge is made by any one of the users and the challenge method is then carried out for each of the enrollments until either a successful

challenge is made or all of the enrollments have been tested without success.

It is also clear that each of the data sets may include a definition of a specific security action so that when a specific match is made between a challenge and an enrollment, the specific action may be carried out in deference to other actions corresponding to the other enrollments.

While the invention has been described with reference to a preferred embodiment, it is to be clearly understood by those skilled in the art that the invention is not limited thereto. Rather, the scope of the invention is to be interpreted only in conjunction with the appended claims.

What is claimed is:

1. A method of comparing a challenge verbal utterance with an enrollment verbal utterance to determine a security action, the method comprising the steps of:

- A) preparing an enrollment data comprising the step of:
  - a) converting an enrollment verbal utterance into a first electrical signal;
  - b) converting the electrical signal into a digital format;
  - c) creating a fast Fourier transform of the signal to produce a frequency domain representation of the signal;
  - d) taking M sample as frames of the signal for each of N frequency channel to form an M by N sample enrollment matrix  $E_i$ ;
  - e) removing all said frames with zero energy content;
  - f) determining if at least X of the M sample have a selected minimum energy content in at least Y of the N frequency channels;
  - g) repeating steps (a) through (f) if (f) fails, and aborting after a selected number of repeating of steps (a) through (f);
  - h) repeating steps (a) through (f) if step (f) succeeds, until a selected number of said sample matrices  $E_i$  have been formed as  $E_1, E_2, E_3, \dots$ ;
  - i) forming the difference between each pair of the enrollment matrices  $E_i$ , as  $(E_1-E_2), (E_1-E_3), (E_2-E_3), \dots$ ;
  - j) comparing each of the individual frames with each other of the frames by moving each said one of the frames forward and backward in time to obtain the best match;
  - k) forming the sum, S, of the differences between each of the pairs of the matrices  $E_i$  as  $S=\{(E_1-E_2)+(E_1-E_3)+(E_2-E_3)+\dots\}$ ;
  - l) accepting S as an enrollment data if  $S<EB$ , where EB is a selected first enrollment criterion, and if each said difference between each pair of the matrices  $E_i$  is greater than ES, ES being a selected second enrollment criterion;
- B) challenging the enrollment data comprising the steps of:
  - a) converting a challenge verbal utterance into a second electrical signal;
  - b) converting the second electrical signal into a digital format;
  - c) creating a second fast Fourier transform of the signal to produce a frequency domain representation of the second signal;
  - d) taking M samples of the signal for each of N frequency channels to form an M by N sample challenge matrix C;
  - e) removing each of said frames having zero energy content;
  - f) determining if at least X of M samples have a selected minimum energy content in at least Y of the N frequency channels;

- g) repeating steps (a) through (f) if (f) fails, and aborting after a selected number of repeats of steps (a) through (f);
  - h) repeating steps (a) through (f) if step (f) succeeds, until a selected number of said sample matrices C has been formed;
  - i) forming the difference between each pair of the enrollment matrices C, as  $(C-E_1), (C-E_2), (C-E_3), \dots$ ;
  - j) comparing each of the individual frames with each other of the frames by moving each said one of the forward and backward in time to obtain the best match;
  - k) forming the sum, S, of the differences between each of the pairs of the matrices C as  $S=\{(C-E_1)+(C-E_2)+(C-E_3)+\dots\}$ ;
  - l) accepting S as an enrollment data is  $S<VB$ , VB being a selected first verification criterion and also if each said difference between each pair of the matrices C is greater than VS, VS being a selected second verification criterion;
- C) Enabling the security action if the challenge verbal utterance is accepted.
2. A method of comparing a challenge verbal utterance with a set of enrollment verbal utterances to determine a security action, the method comprising the steps of:
- A) preparing enrollment data comprising the steps of:
    - a) converting an enrollment verbal utterance into a first electrical signal;
    - b) converting the electrical signal into a digital format;
    - c) creating a fast Fourier transform of the signal to produce a frequency domain representation of the signal;
    - d) taking M samples as frames of the signal for each of N frequency channels to form an M by N sample enrollment matrix  $E_i$ ;
    - e) removing any said frames having zero energy content;
    - f) determining if at least X of M samples have a selected minimum energy content in at least Y of the N frequency channels;
    - g) repeating steps (a) through (f) if (f) fails, and aborting after a selected number of repeats of steps (a) through (f);
    - h) repeating steps (a) through (f) if step (f) succeeds, until a selected number of said sample matrices  $E_i$  have been formed as  $E_1, E_2, E_3, \dots$ ;
    - i) forming the difference between each pair of the enrollment matrices  $E_i$ , as  $(E_1-E_2), (E_1-E_3), (E_2-E_3), \dots$ ;
    - j) comparing each of the individual frames with each other of the frames by moving each said one of the forward and backward in time to obtain the best match;
    - k) forming the sum, S, of the differences between each of the pairs of the matrices  $E_i$  as  $S=\{(E_1-E_2)+(E_1-E_3)+(E_2-E_3)+\dots\}$ ;
    - l) accepting S as an enrollment data is  $S<EB$ , EB being a selected first enrollment criterion and also if each said difference between each pair of the matrices  $E_i$  is greater than ES, ES being a selected second enrollment criterion
    - m) repeating steps (a) to (l) so as to create a plurality of individual enrollment data sets;
  - B) challenging the enrollment data, the enrollment challenging comprising the steps of:
    - a) converting a challenge verbal utterance into a second electrical signal;

7

- b) converting the second electrical signal into a digital format;
- c) creating a second fast Fourier transform of the signal to produce a frequency domain representation of the second signal;
- d) taking M samples as frames of the signal for each of N frequency channels to form an M by N sample challenge matrix C;
- e) removing any said frames having zero energy content;
- f) determining if at least X of M samples have a selected minimum energy content in at least Y of the N frequency channels;
- g) repeating steps (a) through (f) if (f) fails, and aborting after a selected number of repeats of steps (a) through (f);
- h) repeating steps (a) through (f) if step (f) succeeds, until a selected number of said sample matrices C has been formed;
- i) forming the difference between each pair of the enrollment matrices C, as (C-E1), (C-E2), (C-E3), . . . ;
- j) comparing each of the individual frames with each other of the frames by moving each said one of the forward and backward in time to obtain the best match;

8

- k) forming the sum, S, of the differences between each of the pairs of the matrices C as  $S = \{(C-E1) + (C-E2) + (C-E3) + \dots\}$
  - l) repeating the steps (a) thru (k) for each of the enrollment data sets until  $S' < VB$ , VB being a selected first verification criterion, and each said difference between each of the pairs C and Ei is greater than VS, VS being a selected second verification criterion;
  - m) aborting the method when step (l) has been completed without the challenge verbal utterance being accepted;
  - C) enabling the security action if the challenge verbal utterance is accepted in step B(l).
3. The method of claim 2 wherein each of the data sets includes data defining a specific security action wherein upon completion of step B(l) a further step is taken to complete said specific security action, said action including the generating of an electrical signal.

\* \* \* \* \*